WORKING PAPER
ALFRED P. SLOAN SCHOOL OF MANAGEMENT

MYTHS ABOUT PEOPLE AT WORK NO. 1: PERFORMANCE APPRAISAL SYSTEMS IDENTIFY THE BEST PERFORMERS

James W. Driscoll*

WP 1198-81 March 17, 1981

MASSACHUSETTS
INSTITUTE OF TECHNOLOGY
50 MEMORIAL DRIVE
CAMBRIDGE, MASSACHUSETTS 02139

MYTHS ABOUT PEOPLE AT WORK
NO. 1:
PERFORMANCE APPRAISAL SYSTEMS
IDENTIFY THE BEST PERFORMERS

James W. Driscoll*

WP 1198-81 March 17, 1981

*Assistant Professor
Industrial Relations Section
Sloan School of Management
M. I. T.
617-253-7998

Comments on this paper are welcome.

In mid July, 1979, President Jimmy Carter came down from the mountain at Camp David and fired several key Cabinet members including Secretary of Health, Education, and Welfare Joe Califano. His administration was in trouble and the Cabinet shakeup was Carter's signal that he was in command and taking action. To extend the shakeup to sub-Cabinet appointees, the White House developed a "staff evaluation" questionnaire for Cabinet members to appraise the performance of their staffs. (1) The most famous item in the questionnaire asked:

"#21 Rate this person's political skills.

1 2 3 4 5 6

naive savvy"

As a performance evaluation system, Hamilton Jordan's questionnaire (the Chief of Staff was the author) differed only in its blatant subjectivity and ad hoc construction from routine management practice in the U.S.

Performance appraisal means the evaluation of individuals as the basis for personnel actions. Who gets a raise, how much, and who gets the next promotion, all hinge on the results of these appraisals for most working people in the U.S.

The trouble with performance appraisal is simple: it doesn't work. There is no evidence that managers can or do make the objective judgments assumed by performance appraisal systems. As two researchers concluded fifteen years ago, "the ratings reflect primarily the personal-social relationships between the supervisor and the subordinate rather than the output of the subordinate in question."(2)

1 Who is Subject to Performance Apraisal?

Most Americans employed by organizations of any moderate size (say more than 400 employees) fall under a formal performance appraisal system. (In smaller organizations, appraisal is explicitly subjective

742331

and conducted at the discretion of the owner or manager.) Formal appraisal systems specify the procedures for judging workers and usually result in a single, written, numerical rating of performance. A 1975 survey by the Bureau of National Affairs, a Washington, D.C. research and publication house, found that 84 per cent of the companies and government agencies belonging to its Personnel Policies Forum (generally conidered to be representative of moderate to large employers in the U.S., both public sector and private) conducted formal performance appraisals of their office workers.(3) Somewhat fewer (54 per cent) appraised their production workers. Therefore, if you work as a secretary, clerk, technician, or professional, you are more likely to fall victim to this myth of performance appraisal than if you work in a machine shop or in the construction trades. The reason for the difference is simple, at least according to B.N.A: labor unions resist performance appraisal. Nine out of ten employers with no regular procedure for evaluation of production workers were unionized.

Managers and supervisors, although their work is difficult to observe and evaluate, usually suffer performance appraisal as well. A 1977 survey by the Conference Board, another management-oriented research firm , this one based in New York City, reported that 74 per cent formally appraised their lower-level managers and 55 per cent formally appraised their "top" management.

Public sector workers are no exception. In 1975, all of the states responding to a survey (45) described formal systems and 39 of the 45 were state-wide.(5) At the federal level, the 1978 Civil Service Reform Act , building on earlier legislative mandates, required formal performance appraisal by all agencies, except " the C.I.A., the Foreign Service, the General Accounting Office, as well as ... individuals appointed by the President.(6)

Unless you work for these federal watchdogs or do production work in a union shop, if you work for an organization of any moderate size in the U.S., public or private, in all likelihood your performance is appraised.

2 What is performance Appraisal?

Despite variations on a theme,the similarities in the practice of performance appraisal across the job levels and industries in these surveys is striking. Once a year, the boss (your immediate supervisor) officially evaluates your performance and records that estimate for use in determining future wage increases and promotions. The written appraisal forms include one of two domains (or both): behavioral traits or work output. To describe these areas, the supervisor either writes a short essay or simply rates your performance with a number, say from 1 to 5.

The variations on this common theme depend mostly on job level. Managers and office workers are more likely to be rated on traits like

"knowledge of work" and "initiative" while production workers get slightly more emphasis on "dependability" and "attendance".(7) Managers are more likely than non-managers to have the appraisal used as personal feedback to improve thier performance and develop their skills. Managers may also be more likely than non-managers to have their work output evaluated in terms of specific, usually cost-related, objectives, for example, a budgeted amount for overtime expenditures. Non-managers more typically receive numerical ratings in terms of their "quantity of work" and"quality of work". However, even at managerial ranks the Conference Board concluded, "it is possible that the conventiaonal rating scale is the most commonly used approach to ... performance appraisal." (8)

Thus , for most workers in organizational America, a simple numerical rating by your supervisor is probably the primary official determinant of future pay and promotion.

3 What Are the Assumptions of Performance Appraisal?

A moment's reflection or reading any analysis on the subject by an industrial/organizational psychologist would reveal a few key assumptions underlying performance appraisal.(9) Simply stated, the systems assume that performance is:

1. objective;

2. individual;

3. personally-controlled;

4. variable across people; and

5. either a single dimension or reducible to a single dimsenstion.

Unfortunately for advocates of performance appraisal, a long tradition of empirical research in industrial and organizational psychology directly contradicts each assumption.

4 Performance is Subjective

In a recent review of the reserarch on performance ratings, Landy and Farr summarized the factors affecting ratings received by people both in real-life organizations and in more tightly controlled experiments.(10) While the amount of evidence on different factors varies, it should be possible to make money by betting that you will get lower ratings of your performance if any of the following conditions holds, regardless of your actual behavior on the job. If you are black, or if your boss is of a different race from yours, or if you are new in your job, or your immediate coworkers are rated poorly,

or and especially, if your sex is not "appropriate" for your job. If you are a woman trying to succeed in a non-traditional job such as manager or construction worker, you'll probably get lower ratings than a man.

Not only do performance ratings depend on who you are, they also depend on who your boss is. You can expect to get lower ratings if your boss is a man or of a different race from you, or low in self-confidence , or "distant" psychologically, or "cognitively complex" or oriented towards production rather than towards the needs of his or her subordinates. While some of these distinctions among supervisors can only be made by psychologists (and indeed may only appeal to such hobbyists), you had better hope that your competition faces similar downward rating tendencies in their boss.

Given the preponderance of evidence for the influence of subjective factors, about the only good news for the assumption of objectivity has come relativelty recently. A few psychologists have carefully created ideal conditions for the observance of performance. For example in one study, two-minute videotapes were shown of bookstackers in a small room. Student volunteers were instructed either to stack a few (3) or a lot (up to 17) of books in the two-minute period. Not having personally seen these tapes, I can only speculate at the foot-dragging in the slow (poor performance)condition and the mad frenzy of the fast (good performance) condition. Under these and similarly contrived conditions, where brief snapshots of people doing obviously well or poorly on some well-defined task, it is possible for observers to distinguish poor from good performance. What relevance such studies have to the complexities of real-life job performance outlined below is unclear. However, even in such ideal conditions, subjective factors like the race and sex of the workers still bias evaluations of performance (11).

The subjective nature of performance is further documented by the inability of different observers of the same person to agree on a rating of performance. For example, if you, your boss, his (or her) boss, and one of your peers were to rate your performance along with several of your coworkers, the most likely result, based on current research, is a nearly random pattern.(12) People rated highly by one observer would be only slightly more likely to be rated high by another. The best agreement among observers typically found in ratings appears when several people in the same hierarchical position do the rating, for example, two supervisors or two coworkers. Even then the agreement rarely accounts for as much as 10 per cent of the differences in ratings. (13)

5 Performance is Collective

The importance of group performance has dominated industrial/ organizational psychology for four decades. Engineers do not design products; engineering teams do . Managers do not make decisions; organizational networks of interested people do. Secretaries do not

support their bosses; a complex of receptionist, word processor, reproductionist, and travel clerk does. Two Harvard Business School professors documented the importance of the work group in highly publicized studies at the Hawthorne plant of Western Electric just outside Chicago.(14) Obviously few people in a work organization turn out a measurable product independently of other people. In addition, the level of output is heavily constrained by peer group pressure, as the Hawthorne experiments showed (much to the initial frustration of the professors who were trying to increase individual productivity). Not only is performance usually best thought of as a group effort, but the relationships among groups critically determines the success or failure of most projects in organizations of any size.(15) Few psychologists today would feel comfortable abstracting individual performance from the web of relationships that characterize modern theories of organizational behavior. Yet that is precisely the task set for superviors by all the performance ratings systems identified in these various surveys of current practice.

6 Performance is Situational

Factors outside the control of each worker constrain performance. Despite the obvious power of situational limitations, most industrial and organizational psychology has focused on individual effort. Only last year (1980), did two researchers at the University of Texas at Dallas explore the impact of situational constraints.(16) Sure enough, people perform worse at their tasks (and are more frustrated) if they have inadequate support in terms of 1) information, 2) tools and equipment, 3) materials and supplies, and 4) preparation time (16). Indeed, the number one complaint in a national survey of workers conducted in 1973 by the University of Michigan was "obstacles to getting the job done."(17) By contrast, the assumption underlying performance appraisal is that people face identical situations with adequate support such that any difference in behavior is a characteristic of the person being appraised.

7 Performance Does Not Vary Across Individuals

People in work organizations sort themselves out to a large extent into jobs which they can perform. (Of course, many are denied access to other higher-paid jobs which they could also perform.) People do apply for jobs and quit them if they don't like them. Managers also devote substantial time and energy, even in the absence of performance appraisal systems, to weeding out people who can't do the work required or who don't fit in for one reason or another. The net result in most work situations is a large number of people who are all performing the work required. This much is common sense.

While there is little empirical researach on variation in job behavior to support the preceding commonsense observations, two repeated findings about performance appraisal are relevant. First,

supervisors typically make very small distinctions between good and poor performers; and second, they rate almost everybody as a good performer.(18)

Rather than take such recurrent patterns as indicating something about organizational reality, psychologists refer to them as "rating errors", first of "central tendency" (most people cluster together), and second, of "leniency" (most ratings are positive).

A variety of different rating instructions and training programs have been developed to correct these "errors". The life cycle of the various rating formats follow a predictable pattern, according to one review: "(a) particular appraisal technique is first widely praised, then used and rejected, and subsequently replaced by a new, more highly praised technique".(19) The most current fad has been "Behaviorally Anchored Rating Scales" which substitute specific descriptions of behavior for the adjective anchors such as "naive" and "savvy" on the Carter Administration questionnaire. For example, "naive" might be replaced by "Can be expected to invite Congressman Tip O'Neil (a devout Irishman) to a meeting on St. Patrick's Day. There is little evidence, according to Landy and Farr, to "justify the increased time invested in BARS developmental procedures."(20) Extensive training can make supervisors give less positive ratings, spread out people across the rating scale, and make more distinctions among dimensions of performance. There is little evidence, however, that such training effects last over time or improve the validity of the ratings as measures of "true" performance.(21)

8 Performance is Multidimensional

Most jobs require a variety of behaviors and rarely do people perform them all with the same degree of competence. A good teacher may do little research; the best mechanic may have an erratic attendance record. Based on her review of the problem of defining criteria of performance, Patricia Cain Smith made the same point in the language of psychologists:"(a)n overwhelming majority of studies involving statistical analyses of sets of criterion measures finds that these analyses rarely yield a single general factor" (22). For example, she cites one researcher who focused exclusively on knowledge and performance in sales. He found four separate factors just for sales and different measures of sales on three of them.(23)

Yet despite the overwhelming evidence identifying many dimensions in performance, raters rarely make distinctions among the various behaviors or results they are asked to appraise. If you are rated high on one dimension, you will probably be rated high on another. The extreme form of this "halo" effect was demonstrated most vividly in a study at West Point. Ratings based only on learning a cadet's name and address predicted how well the performance of the same cadet would be rated 14 weeks later.(25) Apparently, the raters formed a first impression of each cadet which subsequently influenced their ratings of all the dimensions of performance.

Even where raters can make distinctions among different dimensions of performance, another problem arises, namely how to combine multiple measures of performance. A low-level manager may have thirty or more measures of his or her department's performance over a year, such as actual exenditures versus budget in several categories, such as output, scrap, turnover, and so forth. Here psychologists have little advice. It is known that assigning different weights to each dimension does little to improve the rating as a predictor of behavior.(26) Yet most psychologists would recommend leaving the manager free to assign those weights to particular dimensions that make the most sense to him or her in any given situation.(27) The potential for such discretion to result in bias is obvious.

9 Court Attack on Performance Appraisal

The Courts have recognized the potential for bias in performance appraisal systems (and any pay raises, promotions, or discharges dependent upon performance appraisal). In Brito vs. Zia, the Fifth Circuit Court of Appeals in Denver held performance appraisals to the same standards as any other test.(28) Alfred Brito, a Spanish-surnamed machinist working for a subcontractor at the Atomic Energy Commission, Los Alamos Laboratories, was laid off based on a performance appraisal by his supervisor. The Court upheld his complaint because the subcontractor, Zia Company, "failed to introduce evidence of the validity of the employee performance appraisal test consisting of empirical data demonstrating that the test was significantly correlated with important elements of work behavior relevant to the jobs for which (Brito was) being evaluataed".

As a test, performance appraisals must be job-related, reliable, and valid. In a series of other cases, the courts have upheld the complaints of workers discharged or denied promotion based on performance appraisals which failed to meet these standards.(29) Based on a review of those cases, three researchers who consult widely for management developed the following description of a performance appraisal system that should (but might not) hold up against a challenge in court.(30) To place their recomendations in perspective, I have juxtaposed the relevant descriptions of current practice from the several surveys cited above. Performance appraisal system, they recommend, should be:

1. Formalized and standardized (most are);

2. Derived from a thorough,formal analysis of all jobs being rated (only one firm in three has such analyses; 50per cent have never analyzed <u>any</u> of their managerial jobs for this purpose;

3. Based on other evaluations besides supervisor's ratings (less than 20 per cent involve more than that of the supervisor);

4. Carried out by trained raters (in only one third of the employers were the raters trained in how to conduct a performance

appraisal interview; less that half train their raters and that usually consists of 2 to 4 hours of general orientation; at least 25 per cent train nobody).

In addition, the consultants recommend that 1) the evaluator have substantial daily contact with the ratee; 2) fixed weights be given to each specific measure in arriving at an overall rating; and 3) opportunities for promotion and transfer be posted and not require the recommendation of the supervisor, and 4) administration and scoring of the performance appraisal be standardized.

Clearly it is the exceptional performance appraisal system which can withstand scrutiny by the courts in the face of a discrimination complaint.

10 Conclusion

Performance appraisal systems do rate a few people higher in any organizational setting and these people do tend to get higher pay and more frequent promotions. However, such ratings require supervisors to make complicated judgments abstracting individual performance from the influence of situational factors and from the contributions of other workers. Moreover, most subordinates probably differ very little in overall performance, in substantial part because each individual will demonstrate a different pattern of strengths and weaknesses on various components of the job. Faced with this impossible task, is it surprising that the personal and social biases of the supervisors substantially influence the selection of the chosen few?

In practice, then, organizations face a dilemma with respect to performance appraisal. Some, probably most, are de-emphasizing appraisal as the basis of personnel actions such as promotion and layoff, and are emphasizing its role as the basis of one-on-one counseling between supervisor and subordinate to improve job skills.(31) For these employers, the question remains as to what new basis replaces subjective appraisals. Ideally, more objective criteria will emerge such as seniority in layoffs (moderated by affirmative action) or the subordinate's expressed preference for a posted promotional opportunity. Realistically, the "old boy network" of personal biases will probably continue to operate without the structure of performance appraisal.

Other employers, probably only a small minority, are undertaking the substantial effort required to bring performance appraisal into line with court requirements.(32) The danger faced by this strategy is obvious. To the extent that performance is group-based, situationally-constrained, and multi-dimensional, then no system of individual performance appraisal will eliminate biases in subjective evaluation. An improved system would have to procede from a different and vastly more complicated set of assumptions.

The implications of inaccurate assumptions about performance

apraisal for workers are much simpler and more compelling. To the extent that workers have any power in the organization, they ought to resist the application of current systems of performance appraisal. Unions for production workers have apparently taken that step already, based on existing surveys.

Where a worker is protected by law against discrimination, he or she can protest any negative action to a local anti-discrimination agency and ultimately to the federal Equal Employment Opportunity Commission or Office of Federal Contract Compliance. Of course, such complaints impose substantial costs on the individual. Beyond the investment of personal time and money, there is every reason to believe that the worker who complains will be the victim of to discrimination by the same work organization in the future. For example, Brito was working for Zia only because he had complained to the E.E.O.C. about discrimination in an earlier layoff. Despite the protection of that agreement between Zia and the E.E.O.C., he was laid off illegally again. His compensatory damages for this repeated injury were only $1, and, despite getting back pay, he was not given his job back again. Some time ago, my colleague at M.I.T. Doug Brown sought to find out what happened to people whom the National Labor Relations Board had ordered reinstated because of employer violations of their rights to unionize. The results were not encouraging: some were, some weren't.(33) (I have found no follow-up studies of complainants to other agencies, but there is little reason to expect different treatment.)

In summary, if you get good ratings in performance appraisal now, then you probably have no complaints. (Of course, you shouldn't delude yourself that your rating has much if anything to do with your behavior on the job.) If you don't get good ratings (or you don't like the imposition of arbitrary authority which they represent), resist. If you're black, female, aged between 40 and 70, Spanish-surnamed, Native-American, veteran of the Vietnam era, or handicapped --and you're willing to suffer the disfavor of your employer, bring your complaint to the government anti-discrimination agencies. If you're not a member of one of those legally-protected groups or fear personal reprisal, call your union or employee association and ask them to challenge the system based on the research summaraized here. If you don't have the strength of a collective bargaining agent to represent you on this job and you 're discharged based on poor performance appraisals, you really should consider organizing a union on your next job.

FOOTNOTES

1. Schram, Martin. "The Shakeup: A Demand for Loyalty." The Boston Globe, Friday, July 20, 1979. p. 8.

2. Stockford, L. and Bissell, H.W. "Establishing a Graphic Rating Scale." In W.E. Fleishman (Ed.) Studies in Personnel and Industrial Psychology(Rev. Ed.) Homewood, Illinois: Dorsey, 1967. As cited in Patricia C. Smith. "Behaviors, Results, and Organizational Effectiveness. In Marvin D. Dunnette (Ed.) Handbook of Industrial and Organizational PsychologyChicago, Rand McNally, 1976, pp. 745-776.

3. Bureau of National Affairs, Inc. "Employee Performance: Evaluation and Control." Personnel Policies Forum Survey No. 108. Washington, D.C. , February, 1975.

4. Lazer, R. I. and Wikstrom, W.S. "Appraising Managerial Performance: Current Practices and Future Directions." N.Y.: Conference Board, 1977.

5. Field, H.S. and W.H. Holley. "Performance Appraisal--An Analysis of State-Wide Practices. Public Personnel Management (May-June) 1975, Vol. 4, No. 3, pp. 145-150.

6. Latham, G.P. and K.N. Wexley. Increasing Productivity Through Performance Appraisa . Reading, Massachusetts: Addison-Wesley, 1981, pp.28-29.

7. Bureau of National Affairs, Inc., op. cit. and Lazer and Wikstrom, op. cit.

8. Lazer and Wikstrom, op. cit. p. 23.

9. McCall, Morgan W. and David L. DeVries. "Appraisal in Context: Clashing With Organizational Realities". Technical Report No. 4. Greensboro, N.C.: Center for Creative Leadership, October, 1977.

10. Landy, F.J. and Farr, J.L. "Performance Rating." Psychological Bulletin, 1980, Vol. 87, pp. 72-107.

11. Schmitt, Neal and Martha Lappin. "Race and Sex as Determinants of Mean and Variance of Performance Ratings" Journal of Applied Psychology, 1980, Vol. 65, No. 4, pp. 428-435.

12. Landy and Farr., op. cit.

13. Kavanagh, Michael J., MacKinney, A.C., and Wolins, L. "Issues in Managerial Performance: Multitrait-Multimethod Analysis of Ratings". Psychological Bulletin, 1971, Vol. 75, pp. 34-49.

14. Roethlisberger, F. J. and Dickson, W.J. Management and the

Worker. Cambridge, Massachusetts: Harvard University, 1939.

15. Likert, Rensis. The Human Organization: Its Management and Values. New York: McGraw-Hill, 1967.

16. Peters, L.H., O'Connor, E.J., and Rudolf, C.J. "The Behavioral and Affective Consequences of Performance Relevant Situational Variables. Organizational Behavior and Human Performance, 1980, Vol. 25, pp. 79-86.

17. Herrick, Neal Q. and Robert P. Quinn. "The Working Conditions Survey as a Source of Social Indicators." Monthly Labor Review (April) 1971, Vol. 94, No. 4, Pp. 15-24.

18. Smith, Patricia C. "Behaviors, Results, and Organizational Effectiveness." In Marvin D. Dunnette (Ed.) Handbook of Industrial and Organizational Psychology. Chicago: Rand-McNally, 1976. Pp. 745-776.

19. DeCotiis, Thomas and Petit, A. "The Performance Appraisal Process: A Model and Some Testable Propositions". Academy of Management Review 1978, Vol. 3, Pp. 635-646.

20. Landy and Farr, op. cit., p.85.

21. Ibid.

22. Smith, op. cit., p. 748.

23. Rush, C.H. Jr. "A Factorial Study of Sales Criteria. Personnel Psychology 1953, Vol. 6, Pp. 9-24.

24. Smith, op. cit.

25. Vielhuber, D.P. and Gothiel, E. "First Impressions and Subsequent Ratings of Performance." Psychological Reports 1965, Vol. 17, P. 916. Cited by Smith, op. cit.

26. Smith, op. cit.

27. Ibid.

28. Brito v. Zia Fair Employment Practice Cases Vol. 5, pp. 1203-1212.

29. Lubben, Gary L., Duane E. Thompson, and Charles R. Klasson. "Performance Appraisal: the Legal Implications of Title VII." Personnel (May-June) 1980, Vo. 57, No. 3, pp. 11-21.

30. Ibid.

31. Edwards, K.A. "Fair Employment and Performance Appraisal: Legal Requirements and Practical Guidelines". In D.L.

DeVries (Chair), " Performance Appraisal and Feedback: Flies in the Ointment." Symposium presented at the Annnual Meeting of the American Psychological Association, Washington, D.C., September, 1976.

32. Latham and Wexley., op. cit.

33. Brown, Douglass V. "The Impact of Some N.L.R.B. Decisions. Pp. 18-26, in Gerald G. Somers (Ed.) Proceedings of the Thirteenth Annual Meeting of the Industrial Relations Research Association, Madison, Wisconsin: Industrial Relations Research Association, 1961.